

## PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2002-182969

(43)Date of publication of application : 28.06.2002

(51)Int.Cl.

G06F 12/00

G06F 13/00

G06F 17/30

(21)Application number : 2000-383624

(71)Applicant : TOSHIBA CORP

(22)Date of filing : 18.12.2000

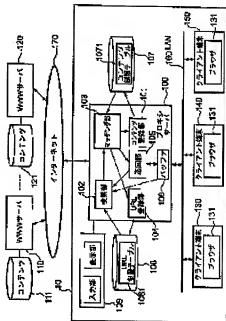
(72)Inventor : SHIMAKAWA KAZUNORI

## (54) PROXY SERVER AND ACCESS LIMITING METHOD

(57)Abstract:

PROBLEM TO BE SOLVED: To efficiently and automatically register an access limited URL to a proxy server and to surely manage contents to be access limited.

SOLUTION: In this proxy server 100, in addition to a URL limitation table 1081, a contents limitation table 1071 for storing access limited contents is provided. Though Internet access is executed by a retrieval part 102 for an access request containing a URL other than the URLs stored in the URL limitation table 1081, in the case that the contents acquired from the Internet are similar to the limited contents registered in the content limitation table 1071, the return of the contents to a client terminal is inhibited. Also, the URL of the acquired contents is automatically added and registered to the URL limitation table 1081 as an access limited address.





## 【特許請求の範囲】

【請求項1】 複数のクライアント端末と外部ネットワークとの間に位置し、前記複数のクライアント端末に代りて前記外部ネットワークへのアクセスを行うプロキシサーバであって、

アクセス制限すべき制限対象コンテンツを記憶するアクセス制限コンテンツ記憶手段と、

前記アクセス制限すべき制限対象コンテンツの存在を示すアドレス情報を記憶するアクセス制限アドレス記憶手段と、

前記各クライアント端末からアクセス要求されたアドレス情報が前記アクセス制限アドレス記憶手段に記憶されているアクセス制限アドレスに該当するか否かを判断し、非該当時に前記アドレス情報を基に前記外部ネットワークにアクセスして、前記アドレス情報で指定されたコンテンツを前記外部ネットワークから取得する検索手段と、

前記検索手段によって前記外部ネットワークから取得されたコンテンツが前記アクセス制限コンテンツ記憶手段に記憶されている制限対象コンテンツに類似しているかどうかを判断し、類似している場合には、アクセス要求元のクライアント端末への前記コンテンツの返送を禁止すると共に、前記コンテンツの存在を示すアドレス情報を前記アクセス制限アドレス記憶手段に追加登録するマッチング処理手段とを具備することを特徴とするプロキシサーバ。

【請求項2】 前記アクセス制限コンテンツ記憶手段には、アクセス制限すべきコンテンツ要素が前記制限対象コンテンツとして記憶されており、

前記マッチング処理手段は、

前記検索手段によって前記外部ネットワークから取得されたコンテンツをコンテンツ要素に分割し、各コンテンツ要素に類似する制限対象コンテンツが前記アクセス制限コンテンツ記憶手段に記憶されているかどうかを判断することを特徴とする請求項1記載のプロキシサーバ。

【請求項3】 前記マッチング処理手段は、

前記アクセス制限コンテンツ記憶手段に記憶されている制限対象コンテンツに類似している取得コンテンツを、前記アクセス制限コンテンツ記憶手段に追加登録する手段をさらに含むことを特徴とする請求項1記載のプロキシサーバ。

【請求項4】 前記マッチング処理手段は、

前記アクセス制限コンテンツ記憶手段に記憶されている制限対象コンテンツに類似しており、且つ前記制限対象コンテンツとの差分が所定値以上の取得コンテンツを、前記アクセス制限コンテンツ記憶手段に追加登録する手段をさらに含むことを特徴とする請求項1記載のプロキシサーバ。

【請求項5】 前記アクセス制限コンテンツ記憶手段に記憶されている制限対象コンテンツと前記マッチング処

理手段によって追加登録されたコンテンツとの差分を算出し、前記制限対象コンテンツとの差分が所定値よりも小さいコンテンツを前記アクセス制限コンテンツ記憶手段から削除する手段をさらに含むことを特徴とする請求項3記載のプロキシサーバ。

【請求項6】 複数のクライアント端末と外部ネットワークとの間に位置し、前記複数のクライアント端末に代りて前記外部ネットワークへのアクセスを行うプロキシサーバであって、

10 アクセス制限すべき制限対象コンテンツを記憶するアクセス制限コンテンツ記憶手段と、

前記アクセス制限すべき制限対象コンテンツの存在を示すアドレス情報を記憶するアクセス制限アドレス記憶手段と、

前記各クライアント端末からアクセス要求されたアドレス情報が前記アクセス制限アドレス記憶手段に記憶されているアクセス制限アドレスに該当するか否かを判断し、非該当時に前記アドレス情報を基に前記外部ネットワークにアクセスして、前記アドレス情報で指定されたコンテンツを前記外部ネットワークから取得する検索手段と、

20 前記検索手段によって前記外部ネットワークから取得されたコンテンツが前記アクセス制限コンテンツ記憶手段に記憶されている制限対象コンテンツに類似しているかどうかを判断し、類似している場合には、アクセス要求元のクライアント端末への前記コンテンツの返送を禁止すると共に、前記コンテンツの存在を示すアドレス情報を前記アクセス制限アドレス記憶手段に追加登録するマッチング処理手段と、

30 前記制限対象コンテンツに類似していると判断された取得コンテンツに関連するコンテンツそれぞれについて前記マッチング処理手段による処理が実行されるように、前記制限対象コンテンツに類似していると判断されたコンテンツに埋め込まれているリンク先アドレス情報を抽出し、その抽出したリンク先アドレス情報に基づいて、前記検索手段に前記外部ネットワークへのアクセスを順次実行させる巡回検索手段とを具備することを特徴とするプロキシサーバ。

【請求項7】 複数のクライアント端末と外部ネットワークとの間に位置するプロキシサーバを用いて、各クライアント端末から外部ネットワークへのアクセスを制限するためのアクセス制限方法であって、

前記各クライアント端末からアクセス要求されたアドレス情報が、アドレス制限テーブルに登録されているアクセス制限アドレスに該当するか否かを判断し非該当時に前記アドレス情報を基に前記外部ネットワークにアクセスして、前記アドレス情報で指定されたコンテンツを前記外部ネットワークから取得する検索ステップと、

50 前記検索ステップによって前記外部ネットワークから取得されたコンテンツが、コンテンツ制限テーブルに登録

3

されている制限対象コンテンツに類似しているかどうかを判別し、類似している場合には、アクセス要求元のクライアント端末への前記コンテンツの返送を禁止すると共に、前記コンテンツの所在を示すアドレス情報を前記アドレス制限テーブルに追加登録するマッチング処理ステップとを具備することを特徴とするアクセス制限方法。

【請求項8】 前記コンテンツ制限テーブルには、アクセス制限すべきコンテンツ要素が前記制限対象コンテンツとして記憶されており、

前記マッチング処理ステップは、前記検索ステップによって前記外部ネットワークから取得されたコンテンツをコンテンツ要素に分割し、各コンテンツ要素に類似する制限対象コンテンツが前記コンテンツ制限テーブルに記憶されているかを否かを判別することを特徴とする請求項7記載のブロキシサーバ。

【請求項9】 前記マッチング処理ステップは、前記コンテンツ制限テーブルに登録されている制限対象コンテンツに類似している取得コンテンツを、前記コンテンツ制限テーブルに追加登録するステップをさらに含むことを特徴とする請求項7記載のアクセス制限方法。

【請求項10】 複数のクライアント端末と外部ネットワークとの間に位置するブロキシサーバを用いて、各クライアント端末から外部ネットワークへのアクセスを制限するためのアクセス制限方法であって、

前記各クライアント端末からアクセス要求されたアドレス情報が、アドレス制限テーブルに登録されているアクセス制限アドレスに該当するか否かを判断し非該当時に前記アドレス情報を基に前記外部ネットワークにアクセスして、前記アドレス情報を指定されたコンテンツを前記外部ネットワークから取得する検索ステップと、

前記検索ステップによって前記外部ネットワークから取得されたコンテンツが、コンテンツ制限テーブルに登録されている制限対象コンテンツに類似しているかどうかを判別し、類似している場合には、アクセス要求元のクライアント端末への前記コンテンツの返送を禁止すると共に、前記コンテンツの所在を示すアドレス情報を前記アドレス制限アドレスとして前記アドレス制限テーブルに追加登録するマッチング処理ステップと、前記制限対象コンテンツに類似していると判断された取得コンテンツに関連するコンテンツそれぞれについて前記マッチング処理ステップが実行されるように、前記制限対象コンテンツに類似していると判断された取得コンテンツに埋め込まれているリンク先アドレス情報を抽出し、その抽出したリンク先アドレス情報に基づいて、前記検索ステップに前記外部ネットワークへのアクセスを順次実行させる巡回検索ステップとを具備することを特徴とするアクセス制限方法。

【発明の詳細な説明】

【0001】

4

【発明の属する技術分野】本発明はブロキシサーバおよびそれを用いたアクセス制限方法に関し、特にインターネットなどの外部ネットワークへのアクセスを制限するためのブロキシサーバおよびそれを用いたアクセス制限方法に関する。

【0002】

【従来の技術】近年、インターネット技術の進展およびオフィス内のネットワーク化に伴い、オフィス内のサーバと、インターネット上に存在する無数のWWW(World Wide Web)サーバを同じ操作で各個人のコンピュータ端末上からアクセスできるイントラネット環境が実現されている。

【0003】しかし、このような環境は便利な反面さまざまな問題を引き起こしている。例えば、業務中にもかかわらず検索関連WWWサイトを閲覧したり、個人的な目的で業務とは無関係のサイトをアクセスしたり、あるいは公序良俗に反するような情報提供を行っているサイトを誰でもアクセスできるといった問題があった。

【0004】この問題に対しては、一般には、ブロキシ(代理)サーバと呼ばれるサーバコンピュータを社内LAN(イントラネット)とインターネットとの間に設置し、このブロキシサーバにアクセス制限URL(Unified Resource Locator)のリストを入手によって登録して、社内から当該URLへのアクセスを禁止するという方法がとられている。

【0005】例えば、文獻(特開平11-242639号公報)には、ブロキシサーバにてアクセス制限URLリストを参照して、WWWサーバへのアクセス可否の判定を行う技術が開示されている。

【0006】

【発明が解決しようとする課題】しかし、従来の方法では、アクセス制限URLリストの登録に人手を介する必要があるため、そのために多くの手間と時間が必要とと共に、爆発的に増加し続けるURLに到底追いつけず、結局はアクセス制限すべきコンテンツの管理が不十分になるという問題があった。

【0007】本発明は上述の事情に鑑みてなされたものであり、アクセス制限URLを効率よく自動登録できるようにし、アクセス制限すべきコンテンツをサーバに管理することが可能なブロキシサーバシステムおよびアクセス制限方法を提供することを目的とする。

【0008】

【課題を解決するための手段】上述の課題を解決するため、本発明は、複数のクライアント端末と外部ネットワークとの間に位置し、前記複数のクライアント端末に代わって前記外部ネットワークへのアクセスを行うブロキシサーバであって、アクセス制限すべき制限対象コンテンツを記憶するアクセス制限コンテンツ記憶手段と、前記アクセス制限すべき制限対象コンテンツの所在を示すアドレス情報を記憶するアクセス制限アドレス記憶手段

と、前記各クライアント端末からアクセス要求されたアドレス情報が前記アクセス制限アドレス記憶手段に記憶されているアクセス制限アドレスに該当するか否かを判断し非該当時に前記アドレス情報を基に前記外部ネットワークにアクセスして、前記アドレス情報で指定されたコンテンツを前記外部ネットワークから取得する検索手段と、前記検索手段によって前記外部ネットワークから取得されたコンテンツが前記アクセス制限コンテンツ記憶手段に記憶されている制限対象コンテンツに類似しているかどうかを判別し、類似している場合には、アクセス要求元のクライアント端末への前記コンテンツの返送を禁止すると共に、前記コンテンツの所在を示すアドレス情報を前記アクセス制限アドレス記憶手段に追加登録するマッピング処理手段とを具備することを特徴とする。

【0009】このプロキシサーバにおいては、アクセス制限コンテンツの所在を示すアドレス情報を記憶するためのアクセス制限アドレス記憶手段に加えて、アクセス制限コンテンツを記憶するためのアクセス制限コンテンツ記憶手段が設けられており、アクセス制限アドレス記憶手段に記憶されているアクセス制限アドレス以外のアドレス情報を含むアクセス要求であっても、そのアドレス情報をもとに外部ネットワークから取得されたコンテンツがアクセス制限コンテンツに類似している場合には、クライアント端末への返送は禁止される。しかも、その類似コンテンツの所在を示すアドレス情報はアクセス制限アドレス記憶手段にアクセス制限アドレスとして自動的に追加登録されるので、以降は、そのアドレス情報を含むアクセス要求は受付られなくなる。よって、例えば、アクセスを禁止したいコンテンツの内容を示す代表的な、画像やテキスト、音楽などのコンテンツ要素をアクセス制限コンテンツ記憶手段に一旦登録しておくだけで、後は自己増殖的にアクセス制限アドレスの自動登録が行われるようになるので、WWWサイトの爆発的な増加に追いついていくことが可能となる。

【0010】類似コンテンツであるかどうかの判別は、外部ネットワークから取得されたコンテンツをコンテンツ要素に分割し、各コンテンツ要素に類似する制限対象コンテンツがアクセス制限コンテンツ記憶手段に記憶されているかどうかを判別することによって実現できる。

【0011】また、外部ネットワークから取得されたコンテンツがアクセス制限コンテンツに類似している場合には、その類似コンテンツの所在を示すアドレス情報をアクセス制限アドレスとして追加登録するだけでなく、その類似コンテンツを新たなアクセス制限コンテンツとして追加登録することが好ましい。これにより、互いに異なる部分を持つ複数の類似コンテンツそれぞれを類似度判定の基準として用いることができるので、アクセス制限したいコンテンツに該当するコンテンツであるか否かを多角的に判断することが可能となる。

【0012】また、制限対象コンテンツに類似していると判断された取得コンテンツに照め込まれているリンク先アドレス情報を抽出し、その抽出したリンク先アドレス情報に基づいて、検索手段に前記外部ネットワークへのアクセスを順次実行させる巡回検索手段をさらに設けることが好ましい。これにより、上述のマッピング手段によるコンテンツの類似度判定が、類似コンテンツに含まれる複数のリンク先アドレス情報をもとて順次実行されるので、自己増殖的にアクセス制限アドレスを増加することが可能となる。

【0013】

【発明の実施の形態】以下、図面を参照して本発明の実施形態を説明する。図1には、本発明の一実施形態に係るプロキシサーバを用いたアクセス制限システムの構成が示されている。本システムは、社内イントラネットシステム10と外部のインターネット170との間にプロキシサーバ100を設け、そのプロキシサーバ100をインターネット170への接続窓口として使用することにより、インターネット170へのアクセス制限を行うものである。

【0014】図1に示されているように、インターネット170上には多数のWWWサーバ110、120…が存在しており、これらWWWサーバ110、120…と社内イントラネットシステム10のプロキシサーバ100とがインターネット170を介して接続されている。また、社内イントラネットシステム10内においては、プロキシサーバ100は、例えばパーソナルコンピュータ等からそれぞれ構成される複数のクライアント端末（処理端末）130、140、150…にLAN等の社内ネットワーク160を介して接続されている。

【0015】WWWサーバ110、120…はそれぞれHTTP(Hypertext Transfer Protocol)サーバとして機能するものであり、提供対象のコンテンツ111、121…を管理している。クライアント端末130、140、150にはHTTPクライアントとして機能するWWWブラウザ131が搭載されており、これら各クライアント端末130、140、150は、図示しないCPUがWWWブラウザ131を実行することにより、インターネット端末(WWWクライアント)として動作する。

【0016】クライアント端末130、140、150のWWWブラウザ131からインターネット170上のコンテンツ111、121、…を指定するURL(Unified Resource Locator)が入力されると、このURLは、プロキシサーバ100に送信され、このプロキシサーバ100を経由して外部のインターネット170にアクセス要求が送信される。WWWサーバ110、120は自身が管理するURLへのアクセス要求だけを受信処理し、該当するコンテンツ111、121を再びインターネット170を介して要求元

7

のプロキシサーバ100に送信する。これが検索結果としてプロキシサーバ100にて受信された後、社内LAN160上の要求元クライアント端末のWWWブラウザ131に返却された画面上に表示される。

【0017】このように、プロキシサーバ100は複数のクライアント端末130、140、150に代行して外部のインターネット170へのアクセスを行うサーバコンピュータであり、クライアント端末130、140、150から見ればWWWサーバとして機能し、インターネット170上のWWWサーバ110、120からみればWWWクライアントとして機能する。

【0018】このプロキシサーバ100には、図示のように、コンテンツ登録部101、検索部102、マッチング部103、URL登録部104、巡回部105、およびバッファ106が設けられている。また、プロキシサーバ100には、記憶装置としての磁気ディスク装置107、108が接続されている。同様にプロキシサーバ100には、キーボードやマウスなどの入力部とディスプレイなどの表示部とを持つ出力装置109が接続されている。

【0019】コンテンツ登録部101は、ユーザによるクライアント端末130、140、150の操作または入出力装置109の操作に基づいて、アクセスを制限すべきコンテンツを登録する機能を持つ。このコンテンツ登録部101によって登録された制限コンテンツ（テキスト、音楽、静止画、動画など）は磁気ディスク装置107のコンテンツ制限テーブル1071に格納される。また、コンテンツ登録部101は、インターネットから取得されたコンテンツのうち、後述するマッチング部103によってコンテンツ制限テーブル1071内の制限コンテンツに類似していると判断されたコンテンツもそれをマッチング部103から受け取ってコンテンツ制限テーブル1071に新規に登録する機能を持ち合わせている。このようにコンテンツ登録部101は制限すべきコンテンツをコンテンツ制限テーブル1071に登録するために用いられる。

【0020】検索部102は、ユーザによるクライアント端末130、140、150の操作または入出力装置109の操作に基づいて発行されるインターネットアクセス要求に応じて、インターネット170上の該当するWWWサーバのコンテンツをアクセスする機能を持つ。この検索部102は、アクセス要求に含まれるURLを基にインターネット170でアクセスし、URLで指定されたコンテンツを該当するWWWサーバ110、120から取得する。このインターネットアクセスに際しては、アクセス要求に含まれるURLが磁気ディスク装置108のURL制限テーブル1081に登録されているかどうかをチェックし、登録されていれば、インターネットアクセスを行わずに、アクセス要求元であるクライアント端末130、140、150またはプロキシサーバ

8

バ100の入出力装置109に対してアクセス制限URLであることを示すメッセージを返す。そうでなければ、インターネットアクセスを行い、インターネット170上から検索・取得したコンテンツをマッチング部103に送る。

【0021】マッチング部103は、パターンマッチングおよびその他の類似度判定方法によって、コンテンツ制限テーブル1071に登録されているアクセス制限コンテンツと検索部102から渡されたコンテンツとの類似度を判定する。この類似度判定処理（マッチング処理）ではインターネット170から取得されたコンテンツがテキスト、画像などのコンテンツ要素に分解され、その各コンテンツ要素と、コンテンツ制限テーブル1071にアクセス制限コンテンツとして登録されている対応する要素との間で類似度が調べられる。例えば、コンテンツ制限テーブル1071にアクセス制限コンテンツとして画像が登録されており、その画像と類似する画像を含むコンテンツがインターネット170から取得された場合には、そのコンテンツはアクセス制限コンテンツであると判定される。

【0022】類似度マッチングの結果、アクセス制限コンテンツであることが判定したら、マッチング部103はこのURLをURL登録部104に送る。また、アクセス制限コンテンツは、コンテンツ登録部101に送られてコンテンツ制限テーブル1071に新たなアクセス制限コンテンツとして追加登録される。類似するコンテンツ同士であっても互いに異なる部分が存在するので、そのようなコンテンツを新たにアクセス制限コンテンツに加えることにより、アクセス制限したコンテンツの内容を複数のコンテンツによって多角的に表現することが可能となる。これにより、実効的なマッチング処理の精度を高めることが可能となる。

【0023】さらに、マッチング部103は、アクセス制限コンテンツであると判定したコンテンツのURL（アクセス対象URL）と一緒に、そのコンテンツに埋め込まれている他のURLへのリンク情報であるリンクURLリストを巡回部105に送る機能を持つ。

【0024】URL登録部104は、磁気ディスク装置108のURL制限テーブル1081にアクセス制限対象のコンテンツの所在を示すURLに登録するためのものであり、マッチング部103から受け取ったURLが磁気ディスク装置108のURL制限テーブル1081に登録済みであるかどうかを調べ、未登録であれば、それを新規に登録する。また、URL登録部104は、入出力装置109の操作に基づいてアクセスを制限すべきURLに登録する機能も有しており、アクセス制限対象のURLを予めURL制限テーブル1081に登録しておくことができる。

【0025】巡回部105は、リンク情報を送ることでより制限コンテンツに類似すると判断されたコンテン

に関連する他のコンテンツそれぞれを巡回検索するためのものであり、マッチング部103から受け取ったアクセス対象URLとリンクURLリストとに基づいて、検索部102に検索要求を送る。検索部102はブラウザ131から受け取ったURLと同様な処理を行

い、検索結果であるコンテンツを同様にマッチング部103に送る。以下、同様な処理をコンテンツ内にリンクが埋め込まれていない状態となるまで深さ優先(Depth First)の探索アルゴリズムを適用して再帰的に実行する。これにより、制限対象コンテンツに類似していると判断された取得コンテンツに関連する他の関連コンテンツそれぞれについてマッチング処理が自動実行されるので、自己増殖的にアクセス制限URLの登録数を増やしていくことができる。

【0027】図2は、WWWサーバ10、120の管理するコンテンツ(URLの示すHTMLドキュメント)の一般的な構成を示したものである。このコンテンツには、図示のように、テキスト、音、静止画、動画、URLリンクなどが埋め込まれている。

【0028】図3は、コンテンツ制限テーブル1071を図示したものである。制限されているコンテンツ要素のタイプとしては、テキスト、音、静止画、動画の4種類に大別される。コンテンツ制限テーブル1071には、制限したいコンテンツ要素をタイプ別に複数登録することができる。検索部102によって取得されたコンテンツはマッチング部103によって構造解析され、テキスト、音、静止画、動画像に分類され、タイプ別にコンテンツ制限テーブル1071に登録コンテンツとの間の類似度マッチングが行われる。

【0029】図4は、URL制限テーブル1081を図示したものである。URL制限テーブル1081には、URL登録部104が登録したアクセス制限対象となるURLが登録されている。

【0030】図5は、上述の巡回検索のための巡回部105によって作成されるURLリンクツリーを図示したものである。トップURLから従属URLが次々とリンクされており、最終のコンテンツにリンクが埋め込まれていなくなった時点で、このURLリンクツリーの形成は終了する。URLリンクツリーは巡回部105が巡回検索の経路をバッファ106上に順次作成していき、巡回検索が終了した時点でバッファ106からクリアされる。

【0031】例えば、URL1のHTMLコンテンツが類似度マッチングによってアクセス制限コンテンツであると判定された場合、まず、URL1をルートとし、U

URL1のHTMLコンテンツに埋め込まれているURL1、12、13を従属URLとするリンクツリーがバッファ106上に形成される。そして、URL1のHTMLコンテンツに対するアクセスが行われ、そのHTMLコンテンツに含まれるURL11、112が、URL11の従属URLとしてリンクツリーに追加される。次に、URL11のHTMLコンテンツに対するアクセスが行われ、そのHTMLコンテンツに含まれるURL111が、URL111の従属URLとしてリンクツリーに追加される。なお、リンクツリーがループを形成していることを判断した場合には、そこで探索を中止するように制御する。

【0032】この後、URL111のHTMLコンテンツのアクセスが行われるが、このHTMLコンテンツには他のコンテンツのURLが含まれていないので、URL111以下の階層に対する処理はこの時点で終了され、URL112以下の階層に対する処理に移る。

【0033】このようにして、リンクツリーを形成しながら、全ての関連コンテンツに対するアクセスが順次実行される。全ての階層についてリンク先を持たない最終コンテンツにたどり着いた時に巡回検索は終了される。なお、ある階層までの検索で巡回検索処理を途中で終了してもよいことはもちろんである。

【0034】次に、図6のプロフローチャートを参照して、プロキシサーバ100を用いて行われるインターネットアクセス制限処理の手順を説明する。

【0035】まず、アクセスを制限すべきコンテンツの事前登録処理が行われる(ステップS601)。この事前登録処理では、アクセスを制限すべきコンテンツを特定するためのテキスト、音、静止画、または動画像などの要素がプロキシサーバ100の入出力装置109からコンテンツ登録部101に送られ、それがコンテンツ登録部101によってコンテンツ制限テーブル1071に記憶される。コンテンツ制限テーブル1071への制限コンテンツの事前登録が終了すると、クライアント端末130、140、150からのアクセス要求受付が可能となる。

【0036】例えばクライアント端末130のWWWブラウザ131からURLを含むアクセス要求が発行されると、そのアクセス要求がプロキシサーバ100にて受け付けられ(ステップS602)、そのURLが検索部102に送られる(ステップS603)。

【0037】検索部102は受け取ったURLがURL制限テーブル1081に登録されているかどうかをチェックして、アクセス可否の判定を行う(ステップS604)。もし同じURLがすでに登録されているば、検索部102はインターネットアクセスは行わず、アクセス要求元のWWWブラウザ131に対してアクセス拒否メッセージを送る(ステップS605)。

【0038】一方、受け取ったURLがURL制限テー

11

ブル1081に未登録であり、アクセス制限URLに該当しないものである場合には、検索部102は、受け取ったURLを基にインターネット170へのアクセスを開始し、URLで指定されたコンテンツを該当するWWWサーバに要求する(ステップS606)。その後、検索部102は、結果のコンテンツを受信すると、それをURLと共にマッチング部103に送る(ステップS607)。

【0039】マッチング部103は、受け取ったコンテンツを構成要素に分解し、それぞれの要素に似たコンテンツがコンテンツ制限テーブル1071にあるかどうかを類似度判定処理によって調べる(ステップS608)。受け取ったコンテンツの全ての要素について類似するものがコンテンツ制限テーブル1071に存在しない場合には(アンマッチ)、アクセス要求元のWWWブラウザに対してアクセス要求に対する応答としてコンテンツが返される(ステップS609)。

【0040】一方、受け取ったコンテンツの要素と類似するものがコンテンツ制限テーブル1071に存在する場合には(マッチ)、マッチング部103は、該当するコンテンツ要素をコンテンツ登録部101に送り、それをコンテンツ制限テーブル1071に登録する(ステップS610)。次いで、マッチング部103は、受け取ったコンテンツに対応するアクセス対象URLをURL登録部104に送り、それをURL制限テーブル1081に登録する(ステップS611)。この場合、URL登録部104ではアクセス対象URLが登録済みであるかどうか調べられ、登録されない場合にのみURL制限テーブル1081への登録が行われる。

【0041】この後、マッチング部103は、アクセス対象URLと、受け取ったコンテンツから抽出したリンクURLリストを巡回部105に送った後(ステップS612)、アクセス要求元のWWWブラウザに対してアクセス拒否メッセージを返す(ステップS613)。

【0042】以上の処理は、WWWブラウザからアクセス要求を受ける度に実行される。

【0043】このように、図6の処理によれば、URL制限テーブル1081に登録されているアクセス制限URL以外のURLを含むアクセス要求であっても、もしそのURLをもとにインターネット170から取得されたコンテンツがコンテンツ制限テーブル1071のアクセス制限コンテンツに類似している場合には、アクセス要求元へのコンテンツの返送は禁止される。しかも、その類似コンテンツのURLはURL制限テーブル1081に自動的に追加登録されるので、以降は、そのURLを含むアクセス要求は受け付けられなくなる。よって、例えば、アクセスを禁止したいコンテンツの内容を示す代表的な、画像やテキスト、音楽などの要素をコンテンツ制限テーブル1071に一旦登録しておくだけで、後は自己増殖的にアクセス制限URLの登録が行われるよう

12

になるので、WWWサイトの爆発的な増加に追随していくことが可能となる。

【0044】次に、図7のフローチャートを参照して、巡回部105の制御の下に実行される巡回検索処理について説明する。この巡回検索処理の実行タイミングは任意であり、図6の処理と同時に並行して行っても良いし、例えば夜間などにまとめて行うようにしても良い。

【0045】巡回部105は、まず、マッチング部103から受け取ったアクセス対象URLとリンクURLリストを基にバッファ106上にURLリンクツリーを作成する(ステップS701)。次いで、巡回部105は、リンクURLを未処理のURLリンクツリーから順次取り出しながら(ステップS702)、URLリンクツリーに未処理のリンクURLがなくなるまで以下の処理を繰り返す。

【0046】すなわち、まず、URLリンクツリーから取得されたURLが巡回部105から検索部102に送られる(ステップS704)。検索部102は受け取ったURLを基にインターネット170へのアクセスを開始し、URLで指定されたコンテンツを該当するWWWサーバに要求する(ステップS705)。なお、検索部102にてアクセス制限URLであるかどうかを図6のステップS604と同様にして調べ、アクセス制限URLに該当する場合には、インターネットアクセスを行わずに、URLリンクツリーから取得された次のURLについての処理に移行するようにしてもよい。

【0047】WWWサーバからコンテンツを受信すると、検索部102は、それをURLと共にマッチング部103に送る(ステップS706)。マッチング部103は、受け取ったコンテンツを構成要素に分解し、それぞれの要素に似たコンテンツがコンテンツ制限テーブル1071にあるかどうかを類似度判定処理によって調べる(ステップS707)。受け取ったコンテンツの要素と類似するものがコンテンツ制限テーブル1071に存在する場合には(マッチ)、マッチング部103は、該当するコンテンツ要素をコンテンツ登録部101に送り、それをコンテンツ制限テーブル1071に登録する(ステップS708)。

【0048】次いで、マッチング部103は、受け取ったコンテンツに対応するアクセス対象URLをURL登録部104に送り、それをURL制限テーブル1081に登録する(ステップS709)。この場合、URL登録部104ではアクセス対象URLが登録済みであるかどうか調べられ、登録されない場合にのみURL制限テーブル1081への登録が行われる。この後、マッチング部103は、アクセス対象URLと、受け取ったコンテンツから抽出したリンクURLリストを巡回部105に送る(ステップS710)。巡回部105は、受け取ったアクセス対象URLとリンクURLリストを用いてURLリンクツリーの追加処理を行う(ステップS7



01)。

【0049】一方、ステップS707にて、受け取ったコンテンツの全ての要素について類似するものがコンテンツ制限テーブル1071に存在しないと判断された場合には(アンマッチ)、ステップS708、709の処理は行わず、マッチング部103は、アクセス対象URLと、受け取ったコンテンツから抽出したリンクURLリストを巡回部105に送る(ステップS710)。巡回部105は、受け取ったアクセス対象URLとリンクURLリストを用いてURLリンクツリーの追加形成を行(ステップS701)。

【0050】このようにして、URLリンクツリーに未処理のURLが無くなるまで処理が繰り返された後(ステップS703のNO)、巡回部105は、バッファ106のURLリンクツリーを削除して巡回検索処理を終了する(ステップS712)。

【0051】なお、図6および図7の処理では、類似コンテンツと判定されたコンテンツについては無条件にコンテンツ制限テーブル1071に追加登録するようにしたが、ほとんど相違点のないコンテンツを追加登録してもマッチング処理の負荷が大きくなるだけであるので、コンテンツ制限テーブル1071への自動登録処理に際しては、「類似するコンテンツのうち、ある程度の相違点を持つコンテンツのみをコンテンツ制限テーブル1071に追加登録する」という条件を付加することが望ましい。この場合の処理例を図8に示す。

【0052】図8は、図6の検索処理におけるステップS607とS610との間、または図7の巡回検索処理におけるステップS706とS710との間で実行されるものである。

【0053】すなわち、マッチング部103は、受け取ったコンテンツを構成要素に分解し、それぞれの要素に似たコンテンツがコンテンツ制限テーブル1071にあるかどうかを類似度判定処理によって調べる(ステップS801)。受け取ったコンテンツの全ての要素について類似するものがコンテンツ制限テーブル1071に存在しない場合には(アンマッチ)、検索処理の場合には図6のステップS609の処理が、また巡回検索処理の場合には図7のステップS710の処理が実行される。

【0054】一方、受け取ったコンテンツの要素と類似するものがコンテンツ制限テーブル1071に存在する場合には(マッチ)、マッチング部103は、類似度判定の結果に基づき、該当するコンテンツ要素について、コンテンツ制限テーブル1071に登録されている対応する制限対象コンテンツ要素との差分の程度を判断する(ステップS802)。この差分が予め決められた値より大きい場合には、マッチング部103は、その該当するコンテンツ要素をコンテンツ登録部101に送り、それをコンテンツ制限テーブル1071に登録する(ステップS803)。次いで、マッチング部103は、受

け取ったコンテンツに対応するアクセス対象URLをURL登録部104に送り、それをURL制限テーブル1081に登録する(ステップS804)。

【0055】一方、差分が予め決められた値以下である場合には、ステップS803の処理はスキップされ、コンテンツ制限テーブル1071への登録は行われない。

【0056】また、このように検索処理および巡回検索処理の中でコンテンツ制限テーブル1071への登録の有無を判断するのではなく、検索処理および巡回検索処理では図6および図7で説明したように無条件に類似コンテンツをコンテンツ制限テーブル1071に登録しておき、後の任意のタイミングで、追加されたコンテンツ要素の中で元の制限コンテンツとの差分が小さいコンテンツをコンテンツ制限テーブル1071から削除しても良い。このようなコンテンツ制限テーブル1071の更新処理の手順を図9に示す。

【0057】すなわち、まず、マッチング部103は、コンテンツ制限テーブル1071に予め各タイプ別に登録されている元の制限コンテンツ要素毎に、コンテンツ制限テーブル1071に新たに登録された同タイプのコンテンツ要素それぞれとの間の類似度を算出する(ステップS901)。次いで、マッチング部103は、各タイプについて、新たに登録されたコンテンツ要素それぞれについて元の制限コンテンツとの差分が予め決められた値以下であるかどうかを判断し(ステップS902)、そうであれば、そのコンテンツ要素をコンテンツ制限テーブル1071から削除する(ステップS903)。

【0058】なお、以上説明したプロキシサーバ100の機能はすべてソフトウェアとして実現することができる。したがって、本実施形態のプロキシサーバ100の機能・手順を含むコンピュータプログラムをCD-ROMなどのコンピュータ読み取り可能な記憶媒体に記憶しておき、そのコンピュータプログラムを社内LANと外部ネットワークとの間に設置されたコンピュータに読み込んで実行させるだけで、本実施形態と同様の効果を実現することができる。

【0059】また、本発明は、上記実施形態に限定されるものではなく、実施段階ではその要旨を逸脱しない範囲で種々に変形することが可能である。更に、上記実施形態には種々の段階の発明が含まれており、開示される複数の構成要件における適宜な組み合わせにより種々の発明が抽出され得る。例えば、実施形態に示される全構成要件から幾つかの構成要件が削除されても、発明が解決しようとする課題の欄で述べた課題が解決でき、発明の効果の欄で述べられている効果が得られる場合には、この構成要件が削除された構成が発明として抽出され得る。

【0060】

【発明の効果】以上説明したように、本発明によれば、

15

アクセス制限URLを効率よく自動登録できるようになり、アクセス制限すべきコンテンツを確実に管理することが可能となる。

#### 【図面の簡単な説明】

【図1】本発明の一実施形態に係るプロキシサーバを用いたシステム全体の構成を示すブロック図。

【図2】同実施形態のシステムでアクセス対象となるHTMLコンテンツの一般的な構成内容を示す図。

【図3】同実施形態のシステムで使用するコンテンツ制限テーブルの登録内容を示す図。

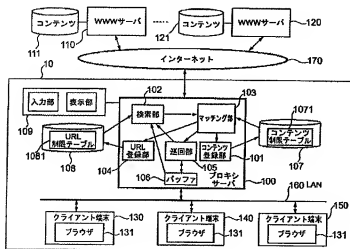
【図4】同実施形態のシステムで使用するURL制限テーブルの登録内容を示す図。

【図5】同実施形態のシステムで使用するURLリンクツリーを説明するための図。

【図6】同実施形態のシステムで実行される検索処理の手順を示すフローチャート。

【図7】同実施形態のシステムで実行される巡回検索処理の手順を示すフローチャート。

【図1】



【図3】

制限URL
http://www.site1.com/aaa.html
http://www.site2.com/bbb.html
http://www.site3.co.jp/ccc.html
⋮
⋮

16

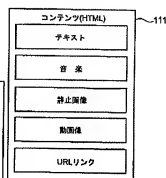
【図8】同実施形態のシステムにおける制限コンテンツ登録処理の手順を示すフローチャート。

【図9】同実施形態のシステムにおけるコンテンツ制限テーブル更新処理の手順を示すフローチャート。

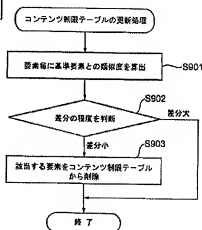
#### 【符号の説明】

- 10…社内イントラネットシステム
- 100…プロキシサーバ
- 101…コンテンツ登録部
- 102…検索部
- 103…マッチング部
- 104…URL登録部
- 105…巡回部
- 110、120…WWWサーバ
- 130、140、150…クライアント端末
- 160…LAN
- 170…インターネット
- 107…コンテンツ制限テーブル
- 108…URL制限テーブル

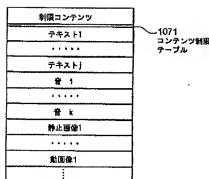
【図2】



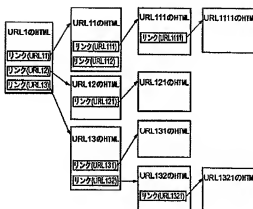
【図9】



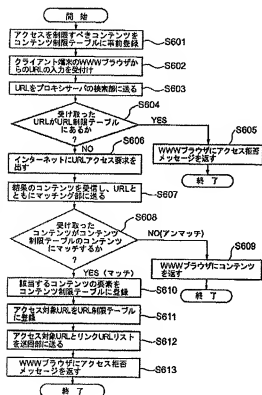
【図 3】



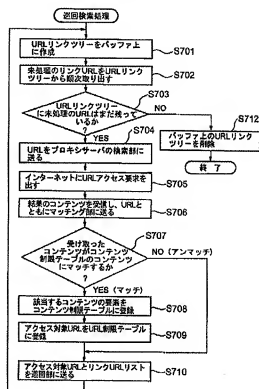
【図 5】



【図 6】



【図 7】



【図 8】

